ECONOMIC RESEARCH REPORTS

The Problem of Prevention

by

Jean-Pierre Benoit
&
Juan Dubra

RR#: 2006-01 **July 2006**

PERSTARE ET PRÆSTARE
MDCCCXXXI

C.V. Starr Center for Applied Economics

Department of Economics

Faculty of Arts and Science

New York University

19 West 4th Street, 6th Floor

New York, New York 10012

The Problem of Prevention*

Jean-Pierre Benoît
Department of Economics and
School of Law, New York University

Juan Dubra
Universidad de Montevideo
Universidad Torcuato di Tella

July, 2006

Abstract

Many disasters are foreshadowed by insufficient preventative care. In this paper, we argue that there is a true *problem of prevention,* in that insufficient care is often the result of rational calculations on the part of agents. We identify three factors which lead to dubious efforts in care. First, when objective risks of a disaster are poorly understood, positive experiences may lead to an underestimation of these risks and a corresponding underinvestment in prevention. Second, redundancies designed for safety, may lead agents to take substandard care. Finally, elected officials have an incentive to underinvest in prevention for some disasters, including those that are relatively unlikely.

Keywords: Prevention, Accidents, Volunteer's Dilemma, Learning, Career Concerns.
Journal of Economic Literature Classification Numbers: D81, D82, D83

A remarkable number of disasters and near-disasters, from the nuclear mishap at Three Mile Island,[1] to the Union Carbide plant tragedy in Bhopal,[2] to the Challenger disaster,[3] to Hurricane Katrina[4] have been preceded by a woefully inadequate level of preventative care, making these adverse events not so much manifestations of poor luck, as all but inevitable occurrences. Indeed, the phrase "an accident waiting to happen" has become somewhat of a cliche in post-event reporting. In this paper, we argue that there is a true *problem of prevention,* in that many accidents are waiting to happen as the result of rational calculations on the part of agents. We identify three factors which lead to dubious efforts in care.

1. When objective risks of a disaster are poorly understood, positive experiences may lead to an underestimation of these risks and a corresponding underinvestment in prevention.

*We thank Martín Besfamille, Federico Echenique, Ennio Staccetti and Federico Weinschelbaum for their comments. Benoît acknowleges the support of the C.V. Starr Center at NYU.

[1] In March 1979, there was a partial meltdown of the reactor core of the Three Mile Island Unit 2 nuclear power plant.

[2] In December 1984, methal isocyanate gas was released at the Union Carbide chemical plant in Bhopal, India, resulting in thousands of deaths and hundreds of thousands of injuries.

[3] The American space shuttle Challenger exploded shortly after takeoff on January 28, 1986.

[4] Hurricane Katrina struck southeast Louisiana on August 29th. Considerable damage was caused, including the flooding of 80% of New Orleans.

2. Redundancies, designed for safety, may lead agents to take substandard care.

3. Elected officials have an incentive to underinvest in prevention for certain disasters, including those that are relatively unlikely.

Some or all of these factors may be present in any given situation. Rather than present a grand model which incorporates all three elements, we present three related models. This permits us to focus sharply on the different effects. Although our models make some contributions of a purely theoretical nature, our primary purpose is of a more applied nature.

Much of the writing on accidents comes from sociologists and psychologists. Vaughan (1996) has written an in-depth study of the Challenger accident in which she faults the "culture" of organizations in general, and of NASA in particular; Perrow (1999) has written about the danger of tightly coupled complex systems, such as Three Mile Island; Reason (1990) has examined the types of errors made by humans, and their causes. We will return to this literature and to the relevant economics literature at various points in this paper.

1 Good News Can Be Bad

The world is a risky place, but how risky is a matter of some choice. Safeguards and redundancies can be built into nuclear power plants, planes can be extensively tested and regularly inspected, space shuttle flights can be cancelled if weather conditions are poor. Just how much effort and expense should be put into preventative care? Among other things, this depends upon the inherent riskiness of the activity involved. But how are the relevant probabilities to be determined?

Scientific and engineering considerations yield a priori probability estimates, which must then be updated in the light of experience. Some industries, such as the airline industry, have a long track record with both successes and failures, so that there is a good understanding of the pertinent probabilities — even when new engines and airplanes are developed, there is a good understanding of the ways in which these need to be tested. Other enterprises, such as nuclear power plants and the space shuttle, involve relatively new technologies with limited experience. These spare histories make it very difficult to estimate the risks involved. In particular, unbroken strings of success make it difficult to assess the probability of a failure. As an example, the space shuttle Challenger had been preceded by twenty-four successful shuttle launches without a failure, and estimates of a catastrophic failure ranged from 1 in 100 to 1 in 100,000 (Feynman (1988))[5]. Similarly, prior to the incident at Three Mile Island there had not been a single accident at a commercial nuclear power plant, and the risks were poorly understood.

Abstractly speaking, any reasonable updating process has the feature that the more time that passes without an adverse incident, the lower the probability that is attached to one. This increasing optimism will lead to a declining investment in precautionary care (under reasonable conditions), and, eventually, to dangerously little care. In this respect, good news can be bad. Investigations into the meltdown at Three Mile Island and the space

[5] It should be noted that the (management) estimate of 1 in $100,000$ is more than a little hard to rationalize.

shuttle Challenger accident showed that such optimistic underinvestment is precisely what took place. With regard to the former, the Kemeny Commission (1979) concluded that:

> "After many years of operation of nuclear power plants, with no evidence that any member of the general public has been hurt, the belief that nuclear power plants are sufficiently safe grew into a conviction. One must recognize this to understand why many key steps that could have prevented the accident at Three Mile Island were not taken. (p.9)."

With regard to the latter, as part of the investigating commission, Feynman (1988)[6] wrote:

> We have also found that certification criteria used in flight readiness reviews often develop a gradually decreasing strictness. The argument that the same risk was flown before without failure is often accepted as an argument for the safety of accepting it again. (p.220)
>
> The *Challenger* flight is an excellent example: there are several references to previous flights; the acceptance and success of these flights are taken as evidence of safety. (p223)
>
> The slow shift toward a decreasing safety factor can be seen in many [areas]. (p230)

Vaughan (1996) has termed this steady decline in standards the "normalization of deviance." We now proceed to a formal model of this phenomenon. To fix our ideas, consider a machine consisting of a single part which may become defective and fail in any period with some fixed unknown probability. In each period, prior to running the machine the part can be tested and, if found defective, costlessly repaired. The test itself, however, is costly and imperfect – at higher costs the test is more likely to detect a defect. We can think of a defective part as an *event*, which turns into an *accident* if and only if it is not detected. With this story in mind, consider the following simple model.

In each period $t = 0, 1, 2...$, nature chooses $y \in \{e, n\}$ (an event occurs or no event occurs) according to some probability $\Pr(y = e) = \hat{\theta} \in (0, 1)$. The parameter $\hat{\theta}$ is unknown, and a decision maker has a belief about $\hat{\theta}$ given by a probability distribution p over $[0, 1]$. For each belief p, the probability of an event is denoted

$$\widehat{p} = \int \theta dp(\theta).$$

An event may or may not turn into an accident. An accident causes a loss of D (in present value terms); the payoff in any single period in which there is no accident is normalized to zero. In each period, the agent can invest in preventative care, which reduces the likelihood of an event turning into an accident. Specifically, the subjective probability of an accident

[6] Feynman's appendix to the commission's report is reprinted in Feynman (1988).

is given by $\phi(\widehat{p},\cdot):\mathbf{R}_{+}\rightarrow[0,\widehat{p}]$, which is continuous and weakly decreasing for all $\widehat{p}$. For an investment c in preventative care, we have[7]:

Probability	**Outcome**
$\phi(\widehat{p},c)$	event and accident
$\widehat{p}-\phi(\widehat{p},c)$	event, but no accident
$1-\widehat{p}$	no event, no accident

There is discounting $\delta\in(0,1)$. Given a prior p, let p_n and p_e denote the (Bayesian) posterior beliefs after no event has happened and after an event has happened, respectively. Starting from a belief p, the Bellman equation for this problem is

$$\begin{aligned} v(p) &= \max_c\left\{\phi(\widehat{p},c)\left[\delta v(p_e)-D\right]+\left[\widehat{p}-\phi(\widehat{p},c)\right]\delta v(p_e)+(1-\widehat{p})\,\delta v(p_n)-c\right\} \\ &= \max_c\left\{-\phi(\widehat{p},c)\,D+\widehat{p}\delta v(p_e)+(1-\widehat{p})\,\delta v(p_n)-c\right\} \end{aligned}$$

In each period, the decision maker's problem is two-fold: to first determine an updated belief p in light of the previous period's experience, and to then choose the optimal c.

The following proposition shows that, quite naturally, the subjective probability that an event will occur falls following no event.

Proposition 1 *For any density p with support $[0,1]$, the probability of an event under beliefs p_n is strictly smaller than under beliefs p. That is, $\widehat{p}_n\equiv\int_0^1\theta p_n(\theta)\,d\theta<\int_0^1\theta p(\theta)\,d\theta=\widehat{p}$.*

Proof. All proofs are in the appendix. ■

Thus, a string of periods with no events leads to a reduced belief in the probability of an event.[8]

Although the Kemeny commission and Feynman appear to take it for granted that increasing optimism leads to declining care, it is easy to think of scenarios in which the reverse is true. Nonetheless, the next proposition supplies plausible conditions under which their intuition is correct.

Proposition 2 *Suppose that ϕ is twice continuously differentiable, with $\phi_{22}>0$ and $\phi_{12}<0$. Then, for each $\hat{p}$, the optimal c, $c(\hat{p})$, is unique. Furthermore, $c(\hat{p}_n)<c(\hat{p})$, for all $\hat{p}$, whenever $c(\hat{p})>0$.*

Propositions 1 and 2 together imply that a string of successes will lead to declining levels of care. If the scientific knowledge is good, so that the a priori estimate of an event is unbiased, then, following the initial period, the level of care will be suboptimal.

[7] The formulation below captures many possibilities. For instance, the probabilities of an event with an accident and of an event without an accident could be of the form $\hat{p}f(c)$ and $\hat{p}(1-f(c))$, respectively, so that $\phi(\widehat{p},c)=\hat{p}f(c)$, or of the form $\hat{p}-f(\hat{p},c)$ and $f(\hat{p},c)$, so that $\phi(\widehat{p},\cdot)=\hat{p}-f(\hat{p},c)$.

[8] Any reasonable updating procedure, not just Bayesian updating, will have the feature that a string of n's leads to an increasing belief in the likelihood of an n, so that the conclusion of Proposition 1 is quite general.

When the potential damage from an accident is very large, the optimal number of accidents is close to zero. For this reason, nuclear reactors are built so that a string of successes is the norm. Unfortunately, our results indicate that this success is to some extent self-defeating. The Kemeny Commission seems to have been aware of this problem, as its report states:

> The Commission is convinced that this attitude [namely, the inference that nuclear plants are safe based on their positive record] must be changed to one that says nuclear power is *by its very nature* potentially dangerous, and, therefore, one must continually question whether the safeguards already in place are sufficient to prevent major accidents (emphasis added). (p9)

Essentially, in its report the commission is imploring nuclear operators to ignore any favorable experience pointing to the safety of nuclear plants. Some of Feynman's recommendations can similarly be interpreted as exhortations to downplay the significance of experience.

At least three factors, two rational and one irrational, make it difficult for agents to ignore, or downplay, experience. First of all, most real world domains are not coin flips with known probabilities, and so general experience must play a role in their evaluation. As reassuring as the Kemeny Commission's recommendation may be, the difficulty with it is readily seen by contemplating a flip side suggestion to insist on the safety of nuclear plants despite a string of disasters. Second of all, there is the role of personal experience. Consider airplane pilots. The presence of idiosyncratic differences among pilots makes it only natural and rational, though perhaps unfortunate, for a particular pilot without an adverse incident to think of himself or herself as particularly skilled, and to be correspondingly less wary than overall probabilities would recommend. Similarly, operators at nuclear power plants may well feel that general experience at plants does not account for the specific conditions at their specific plants. Third of all, and less rationally, as the availability heuristic teaches us, when estimating probabilities people place undue weight on factors that they can readily recall, chief among these being their own experience.

Airplanes are well understood, not only because of their long experience, but also because they are built "bottom up." In contrast with conventional aircraft, the space shuttle was built with a "top down" approach (Feynman (1988)), making it difficult to obtain a tight estimate of the safety of its novel technology. As one might expect, a diffuse prior is relatively prone to learning, so that an observation of a good outcome affects the posterior of a diffuse prior more than a concentrated prior, and, in turn, reduces the level of care of a diffuse prior more than a concentrated prior. Proposition 3 below makes this precise.

Recall that a prior p *second order stochastically dominates* a prior q if they have the same mean and p is better than q for all concave u:

$$p \succeq_s q \Leftrightarrow \int_0^1 z dP(z) = \int_0^1 z dQ(z) \quad \text{and} \quad \int_0^1 u(z)\, dP(z) \geq \int_0^1 u(z)\, dQ(z).$$

Let p_{n^t} be the (Bayesian) posterior of p after t observations of n, and recall that $\widehat{p}_{n^t}$ is the estimated probability of an event based on the distribution p_{n^t}.

Proposition 3 *If* $p \succeq_s q$ *and* $P \neq Q$*, then* $\widehat{p}_{n^t} > \widehat{q}_{n^t}$*. If* $\Phi_{22} > 0$*,* $\Phi_{12} < 0$*, and* $c(p_{n^t}) > 0$*, then,*

$$c(q_{n^t}) < c(p_{n^t})$$

for all t*.*

Thus, innovative technologies are especially susceptible to good news being bad.

We turn now to some related literature.

Our model points to the interaction between learning and investment. As is well understood, for static problems it does not matter whether agents know the probability of an accident, or whether they merely have a distribution of probabilities. When the problem of prevention is repeated over time, however, learning and care-taking interact in non-trivial ways. Gollier (2002) has studied how the curvature (and higher derivatives) of the utility function of the decision maker affect the optimal initial level of care taken when the probability of the accident is unknown. In contrast, our main concern is the study of the evolution of beliefs and how this evolution affects investment over time.

One of the main features of our model, that strings of successes lead to lower care, is reminiscent of the search literature when the distribution that generates wage offers is unknown. This literature has shown that as time goes by, a worker who keeps receiving bad offers becomes more pessimistic about his prospects of finding a decent paying job. He then reduces his reservation wage. The first papers to analyze the decline in reservation wages were, under different assumptions, Rothschild (1974) and Burdett and Vishwanath (1984). Dubra (2004) studies the consequences of this decline in the welfare of the decision-maker.

2 Complex Systems

A lifeguard must continually scan a pool, or a beach, for signs of swimmers in distress. Unfortunately, even highly trained lifeguards may fail to maintain the necessary vigilance.[9] The model that we presented in the previous section suggests that lifeguards who face few emergencies will be especially prone to lapses in vigilance. This finding is consistent with experimental work in psychology which shows that subjects engaged in vigilance tasks perform relatively poorly when the signal rate is low.[10]

While the meandering mind of a lifeguard may prove lethal, the danger posed pales in comparison to the potential harm from a nuclear or chemical plant. For this reason, these plants are designed so that the complacency of a single individual is not sufficient for a disaster to ensue. Consider the following description of an incident at a Union Carbide plant in Institute, West Virginia (Perrow (1999)):

> "[Dangerous] Aldicarb oxime... was transferred to a standby tank that was being pressed into service because of some other problems. Unfortunately, the operators did not know that this tank had a heating blanket and that it was set to come

[9] A 2001 Jeff Ellis & Associates study conducted at 500 swimming pools found that only 9% of lifeguards spotted a submerged mannequin within 10 seconds (considered crucial), and only 43% within 30 seconds.

[10] These vigilance tasks typically last no more than two hours, so that these experiments are not, of course, full blown tests of our theory. See Parasuraman (1981) for a survey.

on as soon as it received product. Also unfortunately, they were not examining the appropriate temperature gauges because they thought there was no need to, and there may have been problems with these anyway because of the nature of the product in the tank. A couple of warning systems failed to activate, and the tank blew... . A few other failures took place..." (p. 358)

Note the number of elements that fell into place to produce this accident: a standby tank was being used *and* there was a heating blanket *and* it was set to come on *and* the operators did not check the temperature gauges *and* warning systems failed *and* the tank blew *and* ... still other things happened. Even with all these failures, there was no loss of life, partly because weather conditions were propitious.

Certainly, the large number of factors that must align in order to produce an accident at a chemical plant contributes to its safety.[11] More generally, consider a system with numerous safety features, all of which must fail for a disaster to result. If the features might fail with given independent probabilities, then the more features, the safer the system. With fully automated features, the logic is unassailable. If humans are involved, however, features that are ostensibly independent may manifest a "strategic dependence," resulting in an ambiguous relationship between reliability and the number of features.

Returning to the Union Carbide case described above, the mere failure of the operators to check the temperature gauges was a long way from producing an accident. But why did the operators fail to check the gauges? The immediate reason given is that "they thought there was no need to," but why did feel no need to follow such an elementary safety precaution?[12] In this section we suggest that at least part of the reason was that the operators knew that even with this lapse, an accident was unlikely, precisely because so many factors had to go awry in order to produce one. That is, the very redundancy features which enhanced the safety of the plant also reduced the incentive of agents to take care, thus limiting the degree of safety that could be achieved.

We turn now to a formal model of this phenomenon.

A disaster may occur. The disaster will happen if and only if each of $n+1$ features fail – an automated feature plus n features under the control of n different people. The probability that the automated feature fails is p_a, while the probability that person i's feature fails is $p(c_i)$, $i = 1, ..., n$, where $c_i \in S = [0, M]$ is the care that i puts in, $p'(c) < 0$, and $p''(c) \geq 0$. Person i's utility function is

$$-p_a \pi_{j=1}^{n} p(c_j) D - c_i,$$

where $D > 0$ reflects the loss of utility from a disaster.

If each person were an automaton simply putting in a designated amount of effort $\bar{c}$, then the probability of a disaster would be $p_a p(\bar{c})^n$. Trivially then, increasing the number n of

[11] Perrow (1999), however, emphasizes the *dynamic* danger of tightly coupled complex systems, such as chemical plants. When things start to go wrong in these systems, it is difficult for workers to understand exactly where the problem lies and how to remedy it on the fly. Thus, whereas we are taking a static view, Perrow is concerned with dynamic difficulties. Nonetheless, Perrow concedes that the number of failures that must take place for an accident to occur, per se, provides a crucial measure of safety.

[12] Similar lapses in care have been noted at numeous other accident sites, including Three Mile Island.

manned features would reduce the probability of an accident, as would better automation in the form of a lower p_a.

Of course, people are not automata; rather, they choose their efforts purposefully. This fact has several consequences. Consider the game in which care levels are chosen simultaneously. As Proposition 4 below indicates, when the number of people increases, each person takes less care in the unique symmetric equilibrium. Similarly, when the automatic feature improves, each person takes less care. An estimation of the safety of the system that neglects strategic slackening will badly miss the mark. While these reductions in care raise the probability of a disaster, increases in the number of people and improvements in automation in and of themselves lower this probability; the net effect is ambiguous. Importantly, under reasonable conditions, increasing the number of people or improving the automated performance may be counterproductive. The following proposition summarizes these findings.

Proposition 4 *The above game has a unique symmetric equilibrium. Let $P(p_a, n)$ be the probability of an accident and $C(p_a, n)$ be the level of care in this equilibrium. Then,*

i) C is decreasing in n
ii) C is increasing in p_a
iii) P may be increasing or decreasing in its arguments.
Suppose that $-p(0)^{n-1} p'(0) > \frac{1}{Dp_a} > -p(M)^{n-1} p'(M)$, so that the equilibrium is interior, and consider $n' > n$ and $p'_a > p_a$. If $\frac{p}{p'}$ is strictly increasing, then $P(p_a, n') > P(p_a, n)$ and $P(p'_a, n) < P(p_a, n)$; if $\frac{p}{p'}$ is strictly decreasing, then $P(p_a, n') > P(p_a, n)$ and $P(p'_a, n) < P(p_a, n)$.

Psychologists have long noted that people working in groups tend to expend less effort than people working as individuals, with larger groups exhibiting more "social loafing."[13] This finding corresponds to i) above. They have also observed that the introduction of automatic devices leads to a decrease in human performance, which corresponds to ii) above.[14] Skitka et al. (2000) put subjects in simulated cockpits with imperfect automated monitoring aids. They then compared the performance of one-person crews with the performance of two-person crews. Although one might naively expect two-person crews to be almost twice as likely to detect system irregularities as one-person crews, they found essentially no difference in detection rates, which is consistent with iii) (albeit in a neutral way).

The following examples illustrate some interesting features of Proposition 4. In the first example, the optimal number of people is an intermediate value.

[13] Pschologists' explanations for social loafing include arousal reduction, decreased evaluation potential, and a matching of anticipated decreased effort on the part of others (see Karau and Williams (1993) for a review).

[14] Psychologists' explanations include automation bias, and automation induced complacency. Consistent with ii), Skitka et. al (1993) find that experimental subjects are less reliable at detecting errors when aided by an automatic system. On the other hand, Parasuraman et al. (1993) conduct an experiment in which they find that the variability in the reliability of an automated system, but not the absolute value of this reliability, affect performance, a finding which is not consitent with ii) (although the interpretation of this finding is confounded by the fact that subjects were not given the reliability parameters).

Example 1 $S = [0,1]\,, p_a D = 40, p(c) = 1 - \frac{5}{4}c + \frac{1}{2}c^2$. *For any* n, *the symmetric equilibrium* c_n *solves* $-p_s D\left(1 - \frac{5}{4}c_n + \frac{1}{2}c_n^2\right)^{n-1}\left(c_n - \frac{5}{4}\right) = 1$.

$$\arg\min_{n} P(p_a, n) = 5$$

In the second example, technological considerations restrict p_a to the interval $\left[\frac{1}{2}, 1\right]$. The probability of an accident $P(p_a, n)$ is minimized by choosing the least reliable automation within this set.

Example 2 $S = [0,1]\,, D > 2, p(c) = (1-c)^b\,,\ 1 \leq b < \frac{n+1}{n},\ p_a \in \left[\frac{1}{2}, 1\right]$. *For any* p_a, *the symmetric equilibrium is* $c = 1 - (bDp_s)^{\frac{1}{1-bn}}$.

$$\arg\min_{p_a \in \left[\frac{1}{2},1\right]} P(p_a, n) = 1$$

Our model is formally a generalization of the Volunteer's Dilemma (Samuelson (1984) and Diekmann (1985)). In this dilemma, an event can be prevented if and only if at least one of n people takes a costly action. Each individual's payoff is given by:

	Someone Else Acts	No One Else Acts
Takes Action	-1	-1
No Action	0	$-D$

Of note, in the symmetric mixed strategy equilibrium the probability of an event is monotonically increasing in n. This model is a special case of Example 2. To see this, set $b = 1, p_a = 1$. Then, a mixed strategy $((1-p), p)$ in the Volunteer's Dilemma corresponds to a pure strategy $c = 1 - p$ in the example. Since the equilibrium is interior, and $\frac{p}{p'} = c - 1$ is an increasing function, (iii) yields the Dilemma result that P is increasing in n. Since Darley and Latané (1968) introduced the concept of "diffusion of responsibility" into the psychology literature, this type of prediction has often been tested, with mixed results (see Goeree, Holt and Moore (2005) and the references therein).

Our results are also reminiscent of the "voluntary provision of public goods" literature. It has long been known that the provision of public goods is subject to a free rider problem, and since Olson (1965) it has been argued that the severity of the problem increases with the number of individuals in society. Since then, several authors have produced examples where the ratio between the optimal amount of a public good and the equilibrium amount of a voluntary provision game increases with the number of players. The only result giving general sufficient conditions for this effect is in Gaube (2001). As in Gaube, we give sufficient conditions for the problem of underprovision to be exacerbated as n increases, but in addition we give sufficient conditions for the amount of the public good provided to be increasing in n. In several other respects, our model is not comparable to this literature. In particular, in voluntary provision models, the public good is the sum of the contributions c_i, whereas in our model it is $(1 - p_a \pi_{i=1}^n p(c_i))$, and the benchmarks used to evaluate the problems are different.

For a complex system, we may expect that, on the one hand, even an operator putting in a minimal amount of effort might detect an anomaly, while on the other hand, even an operator putting in a maximal amount might fail to. Formally, this means $0 < p(M) < p(0) < 1$. Since $p(0) < 1$, the accident-minimizing number of people is then infinity. In practice, however, the "optimal" number of people will be less than infinity, for both technological reasons and economic reasons. As this section emphasizes, there may well be a non-monotonic relationship between the number of people and and the probability of an accident, so that the optimal number of people is not necessarily the "constrained largest." At the same time, since $p(M) > 0$, the optimal number of people is unlikely to be one, in contrast with the Volunteer's Dilemma.

3 Elected Officials

As of this writing, governments throughout the world face the question of how best to deal with the menace of Aviary flu. While experts weigh in with divergent opinions on the danger posed, and by implication the appropriate government action, there is an additional aspect to the problem. In deciding how much to invest in precautionary care, elected officials subject to reelection must consider how their actions will be interpreted by the electorate. In this section, we show that this added concern may cause them to misinvest, even when they have a very good understanding of the threats facing the public. In particular, the officials will have an incentive to underinvest in prevention for potential disasters with relatively low probabilities of occurrence.

Consider the following model. In any period an adverse event may occur with probability p_i, where $i = l$ or h, and $0 \leq p_l < p_h \leq 1$. The prior probability that $p_i = p_h$ is p. An incumbent official, who may or may not be competent, must invest in preventive care. The official has been elected at random, and the electorate initially believes that he, as well as any future candidate, is competent with probability $0 < q < 1$. At the same time, an official believes that he himself is competent with probability $0 < Q \leq 1$. These probabilities are assumed to be common knowledge. Note that if $q \neq Q$ the beliefs of the official and the public are inconsistent;[15] this causes no modeling problems and, when $Q > q$ captures the oft-seen case of officials confident in their own abilities. Whether competent or incompetent, the official receives a private signal $s \in \{h, l\}$. If he is competent, then $s = i$; otherwise $s = h$ with probability p. With this signalling structure, an official's signal is uninformative about his own competence, and an incompetent official's signal is uninformative about the true probability of a disaster.[16]

An official is initially elected for τ periods and may be re-elected exactly once. Each period, he chooses a level of care. The optimal level of care in period t depends upon a priori information, whether or not there have been disasters in periods before t, and the official's

[15] An alternate "consistent" model would have the officials receive private signals regarding their competence. Similar results would obtain. We prefer the model in the text for several reasons, one of which is that it obviates the need for mixed strategies. Of course, the present model can be made consistent by setting $q = Q$.

[16] That is, $p(\text{competent} \mid s) = \frac{\Pr(s|c)\Pr(c)}{\Pr(s|c)\Pr(c) + \Pr(s|i)\Pr(i)} = Q$, and for an incompetent official $p(p_i = p_h \mid s) = p$.

signal s. The only piece of information which is private is the official's signal, and so we can simplify by boiling the official's action down to a declaration of his signal. A strategy for the official is a function $\sigma : \{h, l\} \to [0, 1]$, which maps his signal into a probability with which he announces that h was observed.

In each period, in addition to investment into disaster preparation, the official makes numerous invisible decisions which are more likely to be correct if the official is competent than incompetent. The public wants competent officials in place, and rationally updates its belief about an incumbent official based upon his strategy, his declaration, and the realized pattern of disasters during his initial τ year term. A newly elected official will be competent with probability q. Therefore, the public will re-elect a first term official if its belief that he is competent is greater than q, and will not re-elect him if its belief is less than q. If its belief is exactly q, the official is reelected with a 50% probability. The official only cares about being re-elected (this extreme assumption highlights the problems that arise).

The efficient outcome is for the official to always truthfully report his signal (and make the concomitant investment in care). Unfortunately, under many conditions this will not be equilibrium behavior.

Suppose that the official is certain that he is competent (i.e., $Q = 1$), and hence is certain that his signal correctly reflects the true probability of a disaster (i.e., $\Pr(p_i = p_s \mid s) = 1$). As intuition suggests, this condition maximizes the official's incentive to tell the truth (see Proposition 8 below). Indeed, if his term was arbitrarily long, he would then simply reveal any signal, since, by the law of large numbers, the realization of disasters during his term would then almost surely conform to his signal,[17] and the public's confidence in his competence would increase. The official's term is not arbitrarily long, however. Suppose that this term is, in fact, relatively short and that the probability of a disaster is, at worst, quite small (i.e., $p_h << 1$). Then, *regardless* of his signal, the official ascribes less than a 50% chance to the occurrence of even one event during his initial term. If no event occurs, the public's posterior belief that $p_i = p_h$ will fall (albeit slightly). Suppose the official receives the signal h. He will not want to reveal this signal, since future happenings will more than likely reduce the electorate's belief that $p_i = p_h$, contravening the signal h, and reducing the electorate's belief in his competence if his signal is public. Thus, there is no efficient equilibrium.

What are the conditions for the existence of an efficient equilibrium? Again consider an official who receives the signal h. He is willing to reveal this signal if he expects this revelation to increase the electorate's confidence in him, that is to say, if he believes that subsequent developments are likely to be indicative of p_h. The greater the number of events, N_E, that occur during his term, the more likely that $p_i = p_h$. There is a threshold f so that if the average number of events, $\frac{N_E}{\tau}$, is above this threshold, the probability that $p_i = p_h$ will rise, whereas if the average number is below, this probability will fall. The official will reveal h if $\Pr\left(\frac{N_E}{\tau} \geq f \mid h\right) \geq \frac{1}{2}$. Note that, given h, the official ascribes the probability

[17] More precisely, the official would believe that the realizations would conform to his signal, and this belief is all that matters.

$\Pr(p_i = p_h \mid h)$ to p_h. Therefore, we have that the official will reveal h if

$$\begin{aligned} & \Pr\left(\frac{N_E}{\tau} \geq f \mid h\right) \\ = \; & \Pr(p_i = p_h \mid h) \Pr\left(\frac{N_E}{\tau} \geq f \mid p_h\right) + \Pr(p_i = p_l \mid h) \Pr\left(\frac{N_E}{\tau} \geq f \mid p_l\right) \stackrel{\cdot}{\geq} \frac{1}{2} \end{aligned}$$

Similar reasoning applies to the revelation of the signal l, as the following proposition shows.[18]

An equilibrium is **efficient** if $\sigma(l) = 1 - \sigma(h)$ and $\sigma(l) \in \{0, 1\}$. An equilibrium is **babbling** if $\sigma(l) = \sigma(h)$.

Proposition 5 *Given any p_l, p_h, p, q, Q, τ, a babbling equilibrium always exists. All equilibria are babbling if*

$$\begin{aligned} & \Pr(p_i = p_h \mid h) \Pr\left(\frac{N_E}{\tau} \geq f \mid p_h\right) + \Pr(p_i = p_l \mid h) \Pr\left(\frac{N_E}{\tau} \geq f \mid p_l\right) < \frac{1}{2} \quad (1) \\ \textit{or} \quad & \Pr(p_i = p_h \mid l) \Pr\left(\frac{N_E}{\tau} \leq f \mid p_h\right) + \Pr(p_i = p_l \mid l) \Pr\left(\frac{N_E}{\tau} \leq f \mid p_l\right) < \frac{1}{2} \quad (2) \end{aligned}$$

where

$$f = \frac{\log\left(\frac{1-p_h}{1-p_l}\right)}{\log\left(\frac{p_l}{p_h}\right) + \log\left(\frac{1-p_h}{1-p_l}\right)}.$$

If both the above inequalities are violated strictly, there are, in addition, efficient equilibria (and no others).

Proposition 5 is a bit technical, but it serves as the basis for the remaining, more applied, propositions. Proposition 6 indicates that there are only babbling equilibria when only relatively small probability events or only relatively large probability events are involved – When p_l and p_h are both small there will be underinvestment; when they are both large there will be overinvestment.[19]

Proposition 6 *Fix $p, q \in (0, 1)$, $Q \leq 1$ and $\tau \geq 1$. For low enough p_h, all equilibria are babbling; for large enough p_l, all equilibria are babbling.*

Corollary 1 below provides an illustration of this proposition.

When terms are long, efficient equilibria exist if and only if the signals are reliable enough.[20]

[18] The proposition is stated for generic p_l and p_h (see appendix for details).

[19] The case of large probabilities must be interpreted with some care. If the event occurs with some regularity, such as a Category 2 hurricane, its likelihood will be well understood, and the model will not fit well. On the other hand, if the threat is rare or unique, such as aviary flu, then even if the possible probabilities are large, they will be poorly understood.

[20] When τ is very large, it is *as if* the true state of the world is revealed. In this respect, our model is a generalization of Ottaviani and Prat (2006). However, and importantly, they have a continuous signaling structure so that our model is *not* truly a generalization of theirs.

Proposition 7 *There exists a $\bar{\tau}$ such that for all $\tau > \bar{\tau}$, an efficient equilibrium exists if and only if* $\Pr(p_i = p_h \mid h) > 1/2 > \Pr(p_i = p_h \mid l)$.

The more confident the official, the more likely that an efficient equilibrium exists.

Proposition 8 *Suppose that for a given $p'_l, p'_h, p', q', Q', \tau'$ there is an efficient equilibrium. Then for all $Q > Q'$ there is also an efficient equilibrium for $p'_l, p'_h, p', q', Q, \tau'$.*

The menace posed by Aviary flu is best viewed as a potentiality which will or will not be realized, rather than an event which may or may not occur in successive periods with i.i.d probabilities. We can capture this in our model by setting the official's term τ to 1. Then, p_i is the probability of an outbreak during the official's initial term. The following corollary to Proposition 5 says that there is no efficient equilibrium when the probability of an outbreak is always less than $\frac{1}{2}$, or always greater than $\frac{1}{2}$.

Corollary 1 *When $\tau = 1$, all equilibria are babbling if $p_h < \frac{1}{2}$ or $p_l > \frac{1}{2}$.*

In our formal modeling, officials make explicit announcements of their signals. In practice, governments often implicitly indicate their beliefs by their investments. Consider preparations for a hurricane. The (average member of the) public is likely to take direct note of these preparations only if a hurricane strikes. Absent a hurricane, the expenditure incurred is indirectly noted in that greater expenditure leaves less money over for other items. It was well-recognized before the Category 5 hurricane Katrina struck, that the levees in New Orleans would be inadequate to withstand a strong hurricane. Furthermore, most experts believed that the chance of such a hurricane was high enough to warrant investing in better levees. Indeed, a review by the Army Corps of Engineers later found that the Corps had 'designed the system to protect New Orleans against a relatively low-strength hurricane... and did not respond to warnings over the years from the National Oceanographic and Atmospheric Administration that a stronger hurricane should have been the standard.' (as reported in the New York Times (2006)). At the same time, however, the historical record showed that the probability of a Category 5 hurricane was only 2.4% during a four year period and 4.7% during an 8 year period.[21] Thus, the government's failure to prepare adequately can be understood as a rational bet that developments would increase the public's confidence in it, inadequate preparations notwithstanding. At the same time, given the magnitude of the potential (and actual) damage, the bet was a poor one from an expected value perspective.[22]

For some potential disasters there will be efficient equilibria, for others only babbling equilibria. When there are only babbling equilibria, the public could be better off committing

[21] The return period for Category 5 hurricanes in the New Orleans area is 165, meaning to say that they return, on average, every 165 years, or that 100/165 have occurred in the last 100 years. If this is the result of independent Bernoulli trials, the most likely probability of a hurricane in a given year is approximately 3/500, and the chances that no hurricane will occur in 4 and 8 year periods are the figures given in the text.

[22] Damage that may occur in the future, such as the potential harm from global warming or decaying infrastructure, may lead governments to underinvest due to a lack of concern for future generations, or an inappropriate discount rate. This should not be confused with the present phenomenon; here, the current population is suffering (in expectation).

to re-electing officials, regardless of their perceived competence, since officials would then have no incentive to hide their signals. With the right discount rates, these commitments could arise as part of an equilibrium.[23] However, these equilibria disappear if we assume that officials must pay even a small cost to obtaining their signals, as they would have no incentive to acquire signals if guaranteed re-election.

The model in this section investigates inefficiencies that arise from the interaction between disaster prevention and career concerns on the part of elected officials. In the seminal Holmstrom (1982, reprinted 1999), a manager whose talent is being judged, fails to optimize over project choice, either because he is risk averse or because a "lemons-type" problem arises, neither of which is the case here. Another difference with our model, is that the talents of Holmstrom's managers affect the relevant probabilities, whereas our officials only evaluate probabilities. Scharfstein and Stein (1990) study the inefficiency that arises when managers observe the same signals about the likelihood of success of projects. In their model, managers tend to herd on the choice of projects, so that the market will not be able to update on their ability. The model is similar to ours in that the skill of the official-manager is at evaluating the likelihood of success, and that they also consider a career concerns model. The mechanism whereby the inefficiency arises is different, however. Ottaviani and Sorensen (2004) consider a model in which experts only care about their reputation for competence, and find that they will not truthfully reveal their signals. In their model, the true state of the world is eventually known, whereas in our setting only an update of the state obtains. For this reason, prior probabilities play a much stronger role in their model than ours. Dasgupta and Prat (2004) show that financial traders with reputational concerns may ignore their private information, leading to information cascades.[24]

4 Conclusion

Though an ounce of prevention may be worth a pound of cure, that ounce is often missing. Inadequate care can be the result of miscalculations and other errors. Thus, many analyses of the Challenger disaster emphasize the increasing pressure to launch brought about by the commercialization of the Space Shuttle. We have shown that imprevention can also be the result of a rational calculus.

5 Appendix

For the proof of Proposition 1 we proceed with a series of simple Lemmas concerning the evolution of beliefs. Although versions of the following lemma are well known (see Wolfstetter (1999) Chapter 4) we will use the strict inequalities in this version of our Lemma.

Lemma 1 *If two densities p' and p are such that p'/p is strictly increasing on their support $[0,1]$, then, for all $x \in (0,1)$, their cumulative distribution functions are such that $P'(x) <$*

[23] Thus efficency could be obtained with respect to the officials' strategies, but not also with respect to only re-electing relatively competent officials.

[24] Other models of career concerns include Celentani and Caruana (2001), where managers get to know their types, and Dewatriapont, Jewitt and Tirole (1999) where effort has a cost.

$P(x)$.

Proof. Let $\overline{x}$ be such that $p'(\overline{x}) = p(\overline{x})$. Then, for all $x \in (0, \overline{x})$ we have $p'(x) < p(x)$ and so $P'(x) < P(x)$. For $x > \overline{x}$, $P'(x) - P(x)$ is increasing in x, since the derivative is strictly positive, and therefore is strictly less than $P'(1) - P(1) = 0$. ■

Lemma 2 *For all densities p with support $[0,1]$, the posterior p_n of p is such that $P_n(\theta \leq x) > P(\theta \leq x)$.*

Proof. By Bayes' Rule, the density of the posterior P_n is

$$p_n(\theta) = \frac{\Pr(n \mid \theta)\Pr(\theta)}{\Pr(n)} = \frac{(1-\theta)\,p(\theta)}{\int_0^1 (1-z)\,p(z)\,dz}$$

so that the likelihood ratio of p and p_n is

$$\frac{p(\theta)}{p_n(\theta)} = \frac{\int_0^1 zp(z)\,dz}{1-\theta}$$

which is strictly increasing in θ, so that by Lemma 0, $P_n(\theta \leq x) > P(\theta \leq x)$. ■

Proof of Proposition 1. Since θ is a strictly increasing function, and, by Lemma 2, p strictly dominates $p \mid n$, the result follows. ■

Proof of Proposition 2. To establish uniqueness, fix any p. Suppose that c' and c are optimal for p, with $c' > c \geq 0$. Given our assumptions about differentiability, c' must satisfy the first order condition $-\phi_2(\widehat{p}, c')\,D = 1$, but then $\phi_{22} > 0$ implies that $-\phi_2(\widehat{p}, c)\,D > 1$, so that c can't also be optimal.

If $c(\widehat{p}_n) = 0$ we are done, so suppose that $c(\widehat{p}_n) > 0$. From the first order conditions, $-\phi_2(\widehat{p}, c(\widehat{p}))\,D = 1 = -\phi_2(\widehat{p}_n, c(\widehat{p}_n))\,D$. Since $\widehat{p}_n < \widehat{p}$, $\phi_{12} < 0$, and $\phi_{22} > 0$, we have that $c(p_n) < c(p)$. ■

Proof of Proposition 3. We only need to establish the first part of the proposition, since the second follows in the same manner as Proposition 2. It will suffice to show that $\Pr_{p_{n^t}}(n) < \Pr_{q_{n^t}}(n)$. Notice that for all $t \in (0,1]$, we have that $p \succeq_s q$, $p \neq q$, $(1-\theta)^{t+1}$ strictly convex and $(1-\theta)^t$ concave imply (recall that if a function f is convex, $-f$ is concave)

$$\left.\begin{array}{c} \int_0^1 (1-\theta)^{t+1}\,dP(\theta) < \int_0^1 (1-\theta)^{t+1}\,dQ(\theta) \\ \int_0^1 (1-\theta)^{t}\,dP(\theta) \geq \int_0^1 (1-\theta)^{t}\,dQ(\theta) \end{array}\right\} \Rightarrow \frac{\int_0^1 (1-\theta)^{t+1}\,dP(\theta)}{\int_0^1 (1-\theta)^{t}\,dP(\theta)} < \frac{\int_0^1 (1-\theta)^{t+1}\,dQ(\theta)}{\int_0^1 (1-\theta)^{t}\,dQ(\theta)} \Leftrightarrow$$

$$\int_0^1 (1-\theta)\,d\left(P \mid n^t\right)(\theta) < \int_0^1 (1-\theta)\,d\left(Q \mid n^t\right)(\theta) \Leftrightarrow \Pr_{p_{n^t}}(n) < \Pr_{q_{n^t}}(n).$$

That is, we have established that for all $t \in (0,1]$ we have

$$\Pr_{p_{n^t}}(n) < \Pr_{q_{n^t}}(n) \Leftrightarrow r(t+1) \equiv \frac{\int_0^1 (1-\theta)^{t+1}\,dP(\theta)}{\int_0^1 (1-\theta)^{t+1}\,dQ(\theta)} < \frac{\int_0^1 (1-\theta)^{t}\,dP(\theta)}{\int_0^1 (1-\theta)^{t}\,dQ(\theta)} = r(t). \quad (3)$$

Using this, we will now show that r is decreasing (this last inequality is not enough) by showing that for all $t \in [0,1]$ we have $r'(t+1) < 0$. Since $r'(t+1)$ equals

$$\frac{\int_0^1 (t+1)(1-\theta)^t \, dQ(\theta) \int_0^1 (1-\theta)^{t+1} \, dP(\theta) - \int_0^1 (t+1)(1-\theta)^t \, dP(\theta) \int_0^1 (1-\theta)^{t+1} \, dQ(\theta)}{\left(\int_0^1 (1-\theta)^t \, dQ(\theta)\right)^2}$$

we obtain that $r'(t+1) < 0$ if and only if

$$r(t) = \frac{\int_0^1 (1-\theta)^t \, dP(\theta)}{\int_0^1 (1-\theta)^t \, dQ(\theta)} > \frac{\int_0^1 (1-\theta)^{t+1} \, dP(\theta)}{\int_0^1 (1-\theta)^{t+1} \, dQ(\theta)} = r(t+1) \tag{4}$$

which follows from (3).

Let us assume that, contrary to what we want to prove, $\Pr_{p_{n^s}}(n) \geq \Pr_{q_{n^s}}(n)$ for some s (which from equation (3) must be greater than 1). This happens if and only if for that same s we have $r(s+1) \geq r(s)$. Given this, and that $r'(t) < 0$ for $t \in [1,2]$, let T be defined by

$$T = \min\{t : r'(x) \geq 0\}.$$

We have that $r'(T) = 0$ if and only if (see the calculations in equation 4) $r(T-1) = r(T)$, which is impossible, since by definition of T, for all $t < T$, $r'(t) < 0$. ■

Proof of Proposition 4. First note that the players' strategy spaces are compact and convex, their utility functions are continuous and concave, and the game is symmetric, so the game has at least one symmetric equilibrium. We now show that there is exactly one symmetric equilibrium.

Suppose that $c = 0$ is a symmetric equilibrium. Then

$$-p_a p(0)^{n-1} p'(0) D - 1 \leq 0$$

We have

$$\begin{aligned} \frac{d}{dx}\left(-p_a p(x)^{n-1} p'(x) D - 1\right) &= \\ -p_a (n-1) p(x)^{n-2} p'(x) p'(x) D - p_a p(x)^{n-1} p''(x) D &< 0, \end{aligned} \tag{5}$$

so that

$$-p_a p(x)^{n-1} p'(x) D - 1 < 0 \ \forall x \neq 0$$

and there can be no other symmetric equilibrium. Similarly if $c = M$, there are no other symmetric equilibria. Therefore, if there is a corner symmetric equilibrium, it is the unique symmetric equilibrium.

If $c = \underline{c}$ is an interior symmetric equilibrium then

$$-p_a p(\underline{c})^{n-1} p'(\underline{c}) D = 1,$$

and, since inequality (5) still holds, there is no other interior symmetric equilibrium.

Proof of i) Suppose that $N > n$. If $C(p_a, n) := c$ is interior, then $-p_a p(c)^{n-1} p'(c) D = 1$. Therefore, $-p_a p(c)^{N-1} p'(c) D < 1$, and inequality (5) implies that $C(p_a, N) < C(p_a, n)$. If

$C(p_a,n)=0$, then $-p_a p(0)^{n-1}p'(0)D\le 1$ and $-p_a p(0)^{N-1}p'(0)D<1$, so that $C(p_a,N)=0$. Finally, if $C(p_a,n)=M$, then necessarily $C(p_a,N)\le C(p_a,n)$.

Proof of ii). Suppose that $q_a>p_a$. If $C(p_a,n):=c$ is interior, then $-p_a p(c)^{n-1}p'(c)D=1$. We have, $-q_a p(c)^{N-1}p'(c)D>1$, and inequality (5) implies that $C(q_a,n)>C(p_a,n)$. If $C(p_a,n)=M$, then $-p_a p(0)^{n-1}p'(0)D\ge 1$ and $-q_a p(0)^{N-1}p'(0)D>1$, so that $C(q_a,n)=0$. Finally, if $C(p_a,n)=0$, then necessarily $C(q_a,n)\ge C(p_a,n)$.

Proof of iii). Suppose $-p(0)^{n-1}p'(0)>\frac{1}{Dp_a}>-p(M)^{n-1}p'(M)$ holds. Then, the first order condition at $c=0$ (when all are playing 0) is

$$-p_a p(0)^{n-1}p'(0)D-1>0$$

so that 0 is not a symmetric equilibrium. Similarly, at M, the first order condition is $-p_a p^{n-1}(M)p'(M)D-1<0$, so that $c=M$ for all players is not an equilibrim.

Fix any $n'>n$ and let $C(p_a,n):=c$ and $C(p_a,n'):=c'$. We now show that $P(p_a,n')>P(p_a,n)$ whenever p/p' is strictly increasing. From the proof of i), $c'<c$. Since c is interior, $p(\cdot)/p'(\cdot)$ strictly increasing implies

$$P(p_a,n)=-\frac{p(c)}{Dp'(c)}<-\frac{p(c')}{Dp'(c')}\le P(p_a,n').$$

The proof for $\frac{p(c')}{p'(c')}>\frac{p(c)}{p'(c)}$ follows similarly.

Now suppose that $p_a'>p_a$ and let $c'<c$ be the corresponding equilibrium efforts. An identical argument establishes the desideratum. ■

We now turn to the proofs of the Propositions in Section 3

An event E is a sequence $\{d_s\}_1^\tau$ for $d_s\in\{y,n\}$ for all s; let N_E be the number of disasters in event E. For each fixed τ, and p_x, $x=h,l$, the probability of E given p_x is

$$\Pr(E\mid p_x)=p_x^{N_E}(1-p_x)^{\tau-N_E} \tag{6}$$

so that

$$\frac{\Pr(E\mid p_h)}{\Pr(E\mid p_l)}=\left(\frac{p_h}{p_l}\right)^{N_E}\left(\frac{1-p_h}{1-p_l}\right)^{\tau-N_E}.$$

Then, for (Lebesgue) almost every combination of p_h and p_l, the ratio above is different from 1 for *all* E (all $N_E\le\tau$). Let $S_\tau\subseteq[0,1]^2$ be the set of p_h and p_l such that $\Pr(E\mid p_h)/\Pr(E\mid p_l)\ne 1$ for all E. We will restrict attention to such combinations of p_h and p_l, only to avoid indifferences that only complicate the algebra.

From equation (6) we know that

$$\frac{\Pr(E\mid p_l)}{\Pr(E\mid p_h)}\le 1\Leftrightarrow N_E\ge\tau\frac{\log\left(\frac{1-p_h}{1-p_l}\right)}{\log\left(\frac{p_l}{p_h}\right)+\log\left(\frac{1-p_h}{1-p_l}\right)}\equiv\tau f(p_h,p_l). \tag{7}$$

Then, since for all E, $\Pr(E\mid p_l)\ne\Pr(E\mid p_h)$ (that is, $(p_h,p_l)\in S_\tau$) we have that

$$\Pr(N_E\ge\tau f(p_h,p_l)\mid p_h)=1-\Pr(N_E\le\tau f(p_h,p_l)\mid p_h) \tag{8}$$

and similarly,

$$\Pr\left(N_E \geq \tau f\left(p_h, p_l\right) \mid p_l\right) = 1 - \Pr\left(N_E \leq \tau f\left(p_h, p_l\right) \mid p_l\right) \tag{9}$$

Proof of Proposition 5. Existence of Babbling Equilibria. If the public believes that $\sigma(l) = \sigma(h) = x$, then it will not update for any announcement or event E,and $\sigma(l) = \sigma(h) = x$ is a best response. This holds for any parameters

All non babbling equilibria are efficient, when (1) and (2) are violated strictly. Fix any proposed equilibrium strategy with $\sigma(l) \neq \sigma(h)$. The public's strategy, given $\sigma(l)$ and $\sigma(h)$ must be to rehire if the probability of the individual being competent given the strategies, the announcement $a = h, l$ and the event E,is larger than q. The probability of the individual being competent given $a = h$ is

$$\begin{aligned} P\left(c \mid \sigma, a = h, E\right) &= \frac{P\left(a = h, E \mid c, \sigma\right) q}{P\left(h, E \mid c, \sigma\right) q + P\left(h, E \mid i, \sigma\right)(1-q)} \\ &= \frac{1}{1 + \frac{\left[\left(p\sigma(h)+\sigma(l)(1-p)\right)\frac{P(E|p_h)}{P(E|p_l)}\frac{p}{1-p}+p\sigma(h)+\sigma(l)(1-p)\right](1-q)}{\left[\sigma(h)\frac{P(E|p_h)}{P(E|p_l)}\frac{p}{1-p}+\sigma(l)\right]q}} \end{aligned}$$

This is greater than q iff

$$\begin{aligned} \frac{\left(p\sigma(h) + \sigma(l)(1-p)\right)\frac{P(E|p_h)}{P(E|p_l)}\frac{p}{1-p} + p\sigma(h) + \sigma(l)(1-p)}{\sigma(h)\frac{P(E|p_h)}{P(E|p_l)}\frac{p}{1-p} + \sigma(l)} &\leq 1 \Leftrightarrow \\ \left(\sigma(h) - \sigma(l)\right)\left[\frac{P\left(E \mid p_h\right)}{P\left(E \mid p_l\right)} - 1\right] &\geq 0. \end{aligned} \tag{10}$$

The probability of being competent if the individual announces $a = l$ is

$$P\left(c \mid \sigma, a = l, E\right) = \frac{1}{1 + \frac{\left[\left(p(1-\sigma(h))+(1-\sigma(l))(1-p)\right)\frac{P(E|p_h)}{P(E|p_l)}\frac{p}{1-p}+p(1-\sigma(h))+(1-\sigma(l))(1-p)\right](1-q)}{\left[(1-\sigma(h))\frac{P(E|p_h)}{P(E|p_l)}\frac{p}{1-p}+1-\sigma(l)\right]q}}$$

which is greater than q iff

$$\begin{aligned} \frac{\left(p\left(1-\sigma(h)\right) + \left(1-\sigma(l)\right)(1-p)\right)\frac{P(E|p_h)}{P(E|p_l)}\frac{p}{1-p} + p\left(1-\sigma(h)\right) + \left(1-\sigma(l)\right)(1-p)}{\left(1-\sigma(h)\right)\frac{P(E|p_h)}{P(E|p_l)}\frac{p}{1-p} + 1 - \sigma(l)} &\leq 1 \Leftrightarrow \\ \left(\sigma(h) - \sigma(l)\right)\left[\frac{P\left(E \mid p_h\right)}{P\left(E \mid p_l\right)} - 1\right] &\leq 0 (11) \end{aligned}$$

Assume that $\sigma(h) > \sigma(l)$ (the opposite case is treated similarly). We will show that $\sigma(h) = 1$, by contradiction. Assume that 1 is strictly violated, and that $\sigma(h) < 1$. For $\sigma(h) < 1$ to be optimal it must be that, given the signal h, reporting l yields a utility weakly larger than reporting h. That is, from inequalities 11 and 10, it must be that

$$\begin{aligned} \Pr\left(\left(\sigma(h) - \sigma(l)\right)\left[\frac{P\left(E \mid p_h\right)}{P\left(E \mid p_l\right)} - 1\right] \leq 0 \mid h\right) &\geq \Pr\left(\left(\sigma(h) - \sigma(l)\right)\left[\frac{P\left(E \mid p_h\right)}{P\left(E \mid p_l\right)} - 1\right] \geq 0 \mid h\right) \Leftrightarrow \\ \Pr\left(\frac{P\left(E \mid p_h\right)}{P\left(E \mid p_l\right)} \leq 1 \mid h\right) &\geq \Pr\left(\frac{P\left(E \mid p_h\right)}{P\left(E \mid p_l\right)} \geq 1 \mid h\right) \end{aligned}$$

Since $(p_l, p_h) \in S_\tau$, $\Pr\left(\frac{P(E|p_h)}{P(E|p_l)} \leq 1 \mid h\right) = 1 - \Pr\left(\frac{P(E|p_h)}{P(E|p_l)} \geq 1 \mid h\right)$, the last inequality becomes

$$\begin{aligned}
\frac{1}{2} &\geq \Pr\left(\frac{P(E \mid p_h)}{P(E \mid p_l)} \geq 1 \mid h\right) \\
&= \Pr\left(\left(\frac{p_h}{p_l}\right)^{N_E}\left(\frac{1-p_h}{1-p_l}\right)^{\tau - N_E} \geq 1 \mid h\right) \\
&= \Pr\left(N_E \geq \tau f(p_h, p_l) \mid h\right) > \frac{1}{2} \text{ (since 1 is strictly violated),}
\end{aligned}$$

A contradiction.

We now show that a contradiction obtains if we assume that 2 is strictly violated and $\sigma(l) > 0$. For $\sigma(l) > 0$ to be optimal, after observing l the utility of reporting h must be at least as large as the utility of reporting l. According to equations 11 and 10 this happens iff

$$\begin{aligned}
\Pr\left((\sigma(h) - \sigma(l))\left[\frac{P(E \mid p_h)}{P(E \mid p_l)} - 1\right] \geq 0 \mid l\right) &\geq \Pr\left((\sigma(h) - \sigma(l))\left[\frac{P(E \mid p_h)}{P(E \mid p_l)} - 1\right] \leq 0 \mid l\right) \Leftrightarrow \\
\frac{1}{2} &\geq \Pr\left(\frac{P(E \mid p_h)}{P(E \mid p_l)} \leq 1 \mid l\right) \Leftrightarrow \\
\frac{1}{2} &\geq \Pr\left(\left(\frac{p_h}{p_l}\right)^{N_E}\left(\frac{1-p_h}{1-p_l}\right)^{\tau - N_E} \leq 1 \mid l\right) \\
&= \Pr\left(N_E \leq \tau f(p_h, p_l) \mid l\right) > \frac{1}{2} \text{(1 strictly violated)}
\end{aligned}$$

We conclude that $1 = \sigma(h)$ and $\sigma(l) = 0$, as was to be shown.

Existence of an efficient equilibrium if equations (1) and (2) are violated strictly. We now show that $1 = \sigma(h)$ and $\sigma(l) = 0$, is part of an equilibrium. We must establish two facts. First, having observed h, the official is better off declaring h than declaring l, assuming either declaration would be believed. This is true, according to equations 11 and 10, iff

$$\begin{aligned}
\Pr\left(\frac{P(E \mid p_h)}{P(E \mid p_l)} \geq 1 \mid h\right) &\geq \Pr\left(\frac{P(E \mid p_h)}{P(E \mid p_l)} \leq 1 \mid h\right) \Leftrightarrow \\
\Pr\left(N_E \geq \tau f(p_h, p_l) \mid h\right) &\geq \frac{1}{2}
\end{aligned}$$

which holds since equation (1) is violated. Similarly, after observing l, announcing l is better than announcing h iff

$$\begin{aligned}
\Pr\left(\frac{P(E \mid p_h)}{P(E \mid p_l)} \leq 1 \mid l\right) &\geq \Pr\left(\frac{P(E \mid p_h)}{P(E \mid p_l)} \geq 1 \mid l\right) \Leftrightarrow \\
\Pr\left(N_E \leq \tau f(p_h, p_l) \mid l\right) &\geq \frac{1}{2}
\end{aligned}$$

which is satisfied because (2) is violated.

If equations (1) or (2) hold, the only equilibria are babbling. Suppose that $\Pr(N_E \geq \tau f(p_h, p_l) \mid h) < 1/2$ and that there is an equilibrium with $\sigma(h) > \sigma(l) \geq 0$. From equation 10 we know that $P(c \mid \sigma, a = h, E) \geq q$ iff

$$(\sigma(h) - \sigma(l)) \left[\frac{P(E \mid p_h)}{P(E \mid p_l)} - 1\right] \geq 0 \Leftrightarrow \frac{P(E \mid p_h)}{P(E \mid p_l)} \geq 1$$

and that $P(c \mid \sigma, a = l, E) \geq q$ iff

$$(\sigma(h) - \sigma(l)) \left[\frac{P(E \mid p_h)}{P(E \mid p_l)} - 1\right] \leq 0 \Leftrightarrow \frac{P(E \mid p_h)}{P(E \mid p_l)} \leq 1$$

If the proposed strategies are an equilibrium, then announcing h after observing h must be weakly better than announcing l, which happens iff

$$\begin{aligned}
\Pr\left(\frac{P(E \mid p_h)}{P(E \mid p_l)} \geq 1 \mid h\right) &\geq \Pr\left(\frac{P(E \mid p_h)}{P(E \mid p_l)} \leq 1 \mid h\right) = 1 - \Pr\left(\frac{P(E \mid p_h)}{P(E \mid p_l)} \geq 1 \mid h\right) \Leftrightarrow \\
\Pr\left(\frac{P(E \mid p_h)}{P(E \mid p_l)} \geq 1 \mid h\right) &\geq \frac{1}{2} \Leftrightarrow \Pr(N_E \geq \tau f(p_h, p_l) \mid h) \geq \frac{1}{2}
\end{aligned}$$

but the opposite is true, so $\sigma(h) > \sigma(l)$ cannot be an equilibrium.

Similarly, assume there is an equilibrium with $\sigma(h) < \sigma(l)$. In this case, $1 - \sigma(h) > 1 - \sigma(l) \geq 0$, which means that announcing l after observing h must be weakly better than announcing h, which happens iff

$$\begin{aligned}
\Pr\left((\sigma(h) - \sigma(l)) \left[\frac{P(E \mid p_h)}{P(E \mid p_l)} - 1\right] \leq 0 \mid h\right) &\geq \Pr\left((\sigma(h) - \sigma(l)) \left[\frac{P(E \mid p_h)}{P(E \mid p_l)} - 1\right] \geq 0 \mid h\right) \Leftrightarrow \\
\Pr\left(\frac{P(E \mid p_h)}{P(E \mid p_l)} \geq 1 \mid h\right) &\geq \Pr\left(\frac{P(E \mid p_h)}{P(E \mid p_l)} \leq 1 \mid h\right) \Leftrightarrow \\
\Pr(N_E \geq \tau f(p_h, p_l) \mid h) &= \Pr\left(\frac{P(E \mid p_h)}{P(E \mid p_l)} \geq 1 \mid h\right) \geq \frac{1}{2}
\end{aligned}$$

which does not hold. The only alternative is $\sigma(h) = \sigma(l)$. ■

Proof of Proposition 6. Notice that for p_h sufficiently close to 0

$$\Pr(N_E = 0 \mid p_h) = (1 - p_h)^\tau$$

is close to 1, and since $\tau f(p_h, p_l) > 0$, $\Pr(N_E \geq \tau f \mid p_h)$ becomes arbitrarily close to 0. Since $\Pr(N_E \geq \tau f \mid p_l) < \Pr(N_E \geq \tau f \mid p_h)$, equation (1) tells us that $\Pr(N_E \geq \tau f(p_h, p_l) \mid h) < 1/2$. ■

Proof of Proposition 7. Since $p_h > f(p_h, p_l) > p_l$, and for large enough τ, we have $\Pr(N_E \geq \tau f(p_h, p_l) \mid p_h) \simeq 1$ and $\Pr(N_E \leq \tau f(p_h, p_l) \mid p_l) \simeq 1$, we then obtain that the sufficient conditions in equations (1) and (2) of Proposition (5) hold if and only if $\Pr(p_i = p_h \mid h) > 1/2 > \Pr(p_i = p_h \mid l)$. ■

Proof of Proposition 8. As Q increases from Q' to Q, we see that since

$$\begin{aligned}
\Pr(p_i = p_h \mid h) &= \frac{\Pr(h \mid p_h)\Pr(p_h)}{\Pr(h \mid p_h)\Pr(p_h) + \Pr(h \mid p_l)\Pr(p_l)} \qquad (12) \\
&= \frac{(Q + p(1-Q))p}{(Q + p(1-Q))p + p(1-Q)(1-p)} = Q + p(1-Q) > p \\
\Pr(p_i = p_h \mid l) &= p(1-Q) < p \qquad (13)
\end{aligned}$$

$\Pr(p_i = p_h \mid h)$ increases and and $\Pr(p_i = p_h \mid l)$ decreases. If an efficient equilibrium exists for Q', this means that both $\Pr(N_E \geq \tau f(p_h, p_l) \mid h) \geq 1/2$ and $\Pr(N_E \leq \tau f(p_h, p_l) \mid l) \geq 1/2$, in which case equations (1) and (2) tells us that inequalities are maintained for Q. ■

References

[1] Burdett, K. and T. Vishwanath (1988): "Declining Reservation Wages," *Review of Economic Studies*, **55**, 655-65.

[2] Darley, J. and B. Latané (1968), "Bystander Intervention in Emergencies: diffusion of responsibility," *Journal of Personality and Social Psychology*, **8**, 377-83.

[3] Dasgupta, A., and Prat, A., (2004), "Asset Price Trading When Traders Care About Reputation, Mimeo

[4] Dewatripont, M., I. Jewitt and J. Tirole, (1999), "The Economics of Career Concerns, Part I: Comparing Information Structures," *Review of Economic Studies*, vol. **66**, issue 1, 183-98.

[5] Diekmann, A., (1985), "Volunteer's Dilemma," *Journal of Conflict Resolution*, **29** (4), 605-10.

[6] Dubra, J., (2004) "Optimism and Overconfidence in Search," *Review of Economic Dynamics*, Volume **73**, Issue 1.

[7] Feynman, R., (1988), "What Do You Care What Other People Think?" W.W. Norton and Company, New York.

[8] Gaube, Thomas, (2001), "Group size and free riding when private and public goods are gross substitutes," Economics Letters **70**, 127-132.

[9] Gollier, Christian (2002), "Optimal Prevention of Unknown Risks: a Dynamic Approach with Learning," mimeo University of Toulouse.

[10] Holmstrom, Bengt (1999), "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, **66**, 169-182.

[11] Karau, S., and Williams, K. (1993), "Social Loafing: A Meta-Analytic Review and Theoretical Integration," Journal of Personality and Social Psychology, Vol. 65, 681-706.

[12] Kemeny, John, Bruce Babbitt, Patrick E. Haggerty, Carolyn Lewis, Paul A. Marks, Cora B. Marrett, Lloyd McBride, Harry C. McPherson, Russell W. Peterson, Thomas H. Pigford, Theodore B. Taylor, Anne D. Trunk (1979), "Report of the President's Commission on the Accident at Three Mile Island." Washington, D.C., Government Printing Office.

[13] Olson, Mankur, (1965), *The Logic of Collective Action*. Cambridge/Mass.: Harvard University Press.

[14] Ottaviani, M., and Sorenson, P., (2006) , "Professional Advice," Journal of Economic Theory, 126, 120-142.

[15] Parasuraman, R., (1981) , "The Psychology of Vigilance," Academic Press, London.

[16] Parasuraman, R., Molloy, R., and Singh, L. (1993), "Performance Consequences of Automation-Induced "Complacency, ""International Journal of Aviation Psychology, 3(1), 1-23.

[17] Perrow, C. (1999), "Normal Accidents," Princeton University Press, Princeton, New Jersey.

[18] Reason, J., (1990), Human Error, Cambridge University Press, Cambridge, United Kingdom

[19] Rothschild, M. (1974): "Searching for the Lowest Price When the Distribution of Prices is Unknown," *Journal of Political Economy*, **82**, No. 4.

[20] Samuelson, W., (1984), Potential Entry and Welfare, Mimeo.

[21] Scharfstein, David and Jeremy Stein (1990), "Herd Behavior and Investment," *American Economic Review* **80**(3), 465-79.

[22] Skitka, L., Mosier, K., and Burdick, M., (1999) - "Does Automation Bias Decision-Making?" International Journal of Human-Computers Studies, 51, 991-1006.

[23] Skitka, L., Mosier, K., and Burdick, M., and Rosenblatt, B., (2000) "Automation Bias and Errors: Are Crews Better Than Individuals?" International Journal of Aviation Psychology, 10(1), 85-97.

[24] Vaughan, D., "The Challenger Launch Decision" (1996), The University of Chicago Press, Chicago.